Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT

ELSEVIER

Journal of Computational and Applied Mathematics 164–165 (2004) 1–10

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Bi-factorial analysis for resolution of GPS equations

Pilar Abad*, José P. Suárez

University of Las Palmas de Gran Canaria, Department of Cartography and Graphic Engineering, Las Palmas 35017, Spain

Received 10 September 2002; received in revised form 5 June 2003

Abstract

A great variety of causes are implied in the resolution of GPS equations systems for obtaining precise point positioning. Among them, the time interval between observations and the number of observations are relevant to know the influence in positioning unknowns (X, Y, Z). In this paper, we perform a statistical study showing the possible influence of both factors in the obtained solution. From the results of this paper we conclude that the X and Y solution are independent of the above factors; however, Z shows a significant effect in the manner that time interval and the number of observations are chosen. The results provide insight into the most appropriate design for solving the equation systems of GPS observations.

Keywords: GPS; Factor analysis; Observations

1. Introductions and definitions

The Global Positioning System (GPS) is based on a constellation of about 24 satellites orbiting the earth at altitudes of approximately 26, 000 km. GPS satellites are high enough to avoid problems associated with land-based systems, yet can provide accurate positioning 24 h a day, anywhere in the world.

Any empirical experiment, as for example in GPS observations, has two types of errors: the *systematic* and the *random* error. The systematic error occurs by known causes and can be avoided. The random error is produced by fluctuations of unknown causes. It cannot be corrected before the experiment but it may be evaluated at the end of the experiment.

GPS determines the distance between a GPS satellite and a GPS receiver by measuring the time it takes a radio signal (the GPS signal) to travel from the satellite to the receiver. If the exact

* Corresponding author.
E-mail addresses: mabad@dcegi.ulpgc.es (P. Abad), jsuarez@dcegi.ulpgc.es (J.P. Suárez).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2003.11.006

Table 1
GPS systematic errors

Source	Uncorrected error level (m)
Ionosphere	0–15
Troposphere	0–15
Measurement noise	0–10
Ephemeris data	1–5
Clock drift	0–1.5
Multipath	0–1

time when the signal was transmitted and the exact time when it was received are known, the signal's travel time can be determined. In order to do this, the satellites and the receivers use very accurate clocks which are synchronized in order to generate the same code at exactly the same time.

The GPS system has been designed to be as accurate as possible. However, errors still occur. Added together, systematic errors can cause a deviation of up to 50 m from the actual GPS receiver position, and these must be removed. There are several sources for these errors, the most significant are given in Table 1.

Mechanistic model: mathematical equations: The simultaneous observation of four or more satellites of known position from an unknown place enables us to build an equation system using the distance formula. There are two kinds of equations depending on the satellite signal: *phase* equations and *pseudo-range* equations. Throughout this work, we will focus on the pseudo-range equation.

The GPS observation equation following a mechanistic model, see [1,6], can be written as follows:

$$P_r^s(t) = \rho_r^s(t) + c\,\mathrm{d}t_r(t) + c\,\mathrm{d}T^s(t) + E_\mathrm{I} + E_\mathrm{T} + \varepsilon, \tag{1}$$

where $\rho_r^s(t)$ is the pseudo-range, $\mathrm{d}t_r(t)$ is the station clock offset from GPS time, $\mathrm{d}T^s(t)$ is the satellite clock offset from GPS time, c is the vacuum speed of light, E_I and E_T are the signal path delay due to ionosphere and troposphere, respectively, and ε is the random error distributed as $N(0,\sigma^2)$.

After removing the systematic errors, the simplified observation equation is

$$P_r^s(t) = \rho_r^s(t) + c\,\mathrm{d}t_r(t) + \varepsilon. \tag{2}$$

The pseudo-range $\rho_r^s(t)$ term in (2) is the geometric range computed as a function of satellite coordinates (x_s, y_s, z_s) and station coordinates (x_r, y_r, z_r). According to this, $\rho_r^s(t)$ can be expressed as follows:

$$\rho_r^s(t) = \sqrt{(x_s - x_r)^2 + (y_s - y_r)^2 + (z_s - z_r)^2}.$$

Linearization of observation Eq. (2) around the a priori parameters $X_0 = (x_0, y_0, z_0, t_0)$ becomes:

$$P_r^s(t) = \rho_0^s + \left(\frac{\partial \rho_r^s}{\partial x_r}\right)_0 \mathrm{d}x_r + \left(\frac{\partial \rho_r^s}{\partial y_r}\right)_0 \mathrm{d}y_r + \left(\frac{\partial \rho_r^s}{\partial z_r}\right)_0 \mathrm{d}z_r + c\,\mathrm{d}t_r, \tag{3}$$

where

$$\mathrm{d}x_r = x_r - x_0; \quad \left(\frac{\partial \rho_r^s}{\partial x_r}\right)_0 = \frac{x_0 - x_s}{\rho_0^s},$$

$$\mathrm{d}y_r = y_r - y_0; \quad \left(\frac{\partial \rho_r^s}{\partial y_r}\right)_0 = \frac{y_0 - y_s}{\rho_0^s},$$

$$\mathrm{d}z_r = z_r - z_0; \quad \left(\frac{\partial \rho_r^s}{\partial z_r}\right)_0 = \frac{z_0 - z_s}{\rho_0^s}.$$

Eq. (3) in matrix form is $A\delta + W - V = 0$, where A is the design matrix, δ is the vector of corrections to the unknown parameters $(\mathrm{d}x_r, \mathrm{d}y_r, \mathrm{d}z_r, \mathrm{d}t_r)$, W denotes the correction to preliminary value ρ_0^s and V is the residual vector. The partial derivatives of the observation equations with respect to the two types of parameters, station position $(\mathrm{d}x_r, \mathrm{d}y_r, \mathrm{d}z_r)$ and station clock $(\mathrm{d}t_r)$, form the design matrix A.

The least-squares solution with a priori weights P and constraints P_{X_0} to the parameters is given by $\delta = -(P_{X_0} + A^{\mathrm{T}}PA)^{-1}A^{\mathrm{T}}PW$, so that the estimated parameters are $\hat{X} = X_0 + \delta$.

2. Perturbation analysis of GPS solutions

A key question to understand how GPS works is how approximately must time be kept in order to achieve a specified accuracy of location. The implicit function theorem has been used in [3] allowing one to approximate the timing accuracy required by the GPS system to locate within a given degree of precision. The main result of the application of such theorem is next summarized. If the times t_i are perturbed then coordinates (x_r, y_r, z_r), [1] will change by approximately.

$$\Delta X \approx \frac{\partial X}{\partial t_1}\Delta t_1 + \frac{\partial X}{\partial t_2}\Delta t_2 + \frac{\partial X}{\partial t_3}\Delta t_3 + \frac{\partial X}{\partial t_4}\Delta t_4, \tag{4}$$

$$\Delta Y \approx \frac{\partial Y}{\partial t_1}\Delta t_1 + \frac{\partial Y}{\partial t_2}\Delta t_2 + \frac{\partial Y}{\partial t_3}\Delta t_3 + \frac{\partial Y}{\partial t_4}\Delta t_4, \tag{5}$$

$$\Delta Z \approx \frac{\partial Z}{\partial t_1}\Delta t_1 + \frac{\partial Z}{\partial t_2}\Delta t_2 + \frac{\partial Z}{\partial t_3}\Delta t_3 + \frac{\partial Z}{\partial t_4}\Delta t_4. \tag{6}$$

If a $|\Delta t_s| < K$ for all s and for a given K, then

$$|\Delta Z| \leqslant \left(\left|\frac{\partial Z}{\partial t_1}\right| + \left|\frac{\partial Z}{\partial t_2}\right| + \left|\frac{\partial Z}{\partial t_3}\right| + \left|\frac{\partial Z}{\partial t_4}\right|\right) K \tag{7}$$

and analogous expressions are derived for X and Y. Note that in (7), if the t_i's are perturbed then it is easily obtained by how much do Z change, and conversely, if Z is needed to be within a specified degree of accuracy, how much error time withstand.

[1] Throughout this paper will be denoted by (X, Y, Z).

Taking a simple numerical example of four satellites based on typical values, see [3], $|\Delta X| \leqslant 3.19 \times 10^9$ K, $|\Delta Y| \leqslant 0.7 \times 10^9$ K and $|\Delta Z| \leqslant 5.07 \times 10^9$ K. We may perturb t_i's by more than 3×10^{-8} s and then it would lead to the errors in location $|\Delta X| \leqslant 90$ m, $|\Delta Y| \leqslant 21$ m and $|\Delta Z| \leqslant 150$ m.

On the other hand, the general error equation for GPS error analysis illustrates how the satellite geometry can seriously affect the solutions. This equation can be obtained beginning with the fundamental measurements and proceeding through analysis of the effect of various error sources. To account for the estimated value of (X, Y, Z) and the estimate error $\Delta E = (\Delta X, \Delta Y, \Delta Z)$, Eq. (2) can be changed to [5]:

$$\Delta E = G^{-1} \Delta \rho_r^s. \tag{8}$$

If $\Delta \rho_r^s \leqslant M$ for all s and for a given M, then

$$\Delta E \leqslant G^{-1} M, \tag{9}$$

where G, the geometry matrix, is constructed from the set of approximated directions of the satellites in an Earth Centered–Earth Fixed (EC–EF) frame. Then, G matrix becomes:

$$\begin{pmatrix} \cos(\alpha_1)\cos(\delta_1) & \sin(\alpha_1)\cos(\delta_1) & \sin(\delta_1) & 1 \\ \cos(\alpha_2)\cos(\delta_2) & \sin(\alpha_2)\cos(\delta_2) & \sin(\delta_2) & 1 \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{10}$$

Considering (9) we take the same numerical example as previously [3]. We perturb M's at the most 10 m, so:

$$\begin{pmatrix} -6.48 & 6.62 & -1.61 & 1.47 \\ -0.31 & 1.25 & 0.40 & 1.17 \\ -8.07 & -9.54 & -3.99 & 2.68 \\ 8.71 & -9.54 & 3.92 & -2.09 \end{pmatrix} * \begin{pmatrix} 6 \\ 5 \\ 3 \\ 8 \end{pmatrix} = \begin{pmatrix} 1.16 \\ 2.39 \\ 7.98 \\ -0.41 \end{pmatrix},$$

where $|\Delta X| \leqslant 1.16$ m, $|\Delta Y| \leqslant 2.39$ m and $|\Delta Z| \leqslant 7.98$ m.

It should be noted as the two previous numerical examples reveal that the times and geometry affect the solution. Moreover, the values obtained for the Z seem to have a greater error relative to the X and Y coordinates.

3. Linear model for the two factor analysis

In this section, we summarize the statistical model we use for the study of the GPS observations: *analysis of variance random effects model* [2]. We briefly explain three concepts: A *factor* is an explanatory variable (A, B, C, ...) studied in an investigation. The different values of a factor are called *levels*. A combination of one level from each factor is a *treatment*. Treatments are applied to experimental units. The measurements of the response variable are used to make comparisons among treatments.

The corresponding linear model is as follows [2]:

$$X_{ijk} = \mu + A_i + B_j + e_{ijk} + (AB)_{ij}, \tag{11}$$

where X_{ijk}, following a distribution $N(\mu_{ij}, \sigma^2)$, denotes the measurement of the response variable for the kth experimental unit exposed to the ith level of factor A and the jth level of factor B and μ is a constant. Factor A has I levels: $i=1,\ldots,I$ and factor B has J levels: $j=1,\ldots,J$. The experimental unit range is $k=1,\ldots,K$ for any treatment. e_{ijk} is the experimental error and is independent $N(0,\sigma^2)$. Factors A and B are said to interact $(AB)_{ij}$ if the effect of factor A depends on the level of factor B, or equivalently, if the effect of factor B depends on the level of factor A.

The null hypotheses we will study are:

$$\sigma^2_{Ai} = 0, \tag{12}$$

$$\sigma^2_{Bj} = 0, \tag{13}$$

$$\sigma^2_{ABij} = 0. \tag{14}$$

4. Empirical model: experiment design

The goal in this section is to study the possible influence of two factors in the obtained solution. The statistical model used for the study is the two factor analysis of variance, univariated, random and equilibrated.

The independent variables are factor A = '*number of observations*' and factor B = '*time interval between observations*'. The dependent variables are the position of the station.

For solving the position and clock error of the station, four satellites observed at one time (1 epoch) are needed; this leads to four equations. For factor A we define three levels: (1) 5 epochs, (2) 10 epochs and (3) 20 epochs. For factor B we also define three levels (1) $\Delta t=30''$, (2) $\Delta t=90''$ and (3) $\Delta t = 300''$.

In any experimental unit, we use six satellites. Each treatment (nine in total) consists of 11 different experimental units corresponding to 11 linear equations systems, see Table 2. Rows in Table 2 represent the different number of epochs and columns are the interval between observations. The time interval studied is 12 h since the orbit of the satellites takes half a day. The data are provided by the European Space Operation Center (ESOC) and are taken from the Maspalomas station MAS1 in Gran Canaria, Spain.

Table 2
Experiment description

Epochs	$\Delta t = 30''$	$\Delta t = 90''$	$\Delta t = 300''$
5	11 systems, 30 equations, 8 unknowns	…	…
10	11 systems, 60 equations, 13 unknowns	…	…
20	11 systems, 120 equations, 23 unknowns	…	…

Table 3
Lilliefors test for normality study of variable Z

Epochs	$\Delta t = 30''$	$\Delta t = 90''$	$\Delta t = 300''$
5	0.1945	0.2035	0.1867
10	0.1879	0.1383	0.1746
20	0.1979	0.1621	0.1643

In order to apply the analysis of the variance model, the following assumptions have to be verified [2]:

(1) Independence of samples.
(2) Normality of samples distribution.
(3) Equal variances.

The independence assumption is verified because we choose random observation during the time of study.

The normality assumption can be tested for *Lilliefors test*. This is a *Kolmogorov-Smirnov* approximation where it is not known the population parameters. Using the statistic for a sample with $k=11$, for a significance level of $\alpha=0.05$, the critical value obtained is 0.2490. Table 3 summarizes the results for the nine samples in the Z variable. Can be noted as the computed values are less than 0.2490 and then the null hypotheses are accepted. Repeating the test for X and Y variables we obtain similar results on the null hypotheses. Based on the application of this test we can state that samples distribution are sufficiently close to a normal distribution.

The equal variances can be tested for both factors thanks to *Cochran test* [2] where the statistical null hypothesis is: there is no difference among any variances, $\sigma_1^2=\sigma_2^2=\sigma_3^2$ for three levels in both factors. For a level of significance 0.05 the critical value for Cochran test for three levels and 33 experimental units is 0.493. The Cochran test computes the ratio $R=\max \sigma_i^2/\sum_{i=1}^{3} \sigma_i^2$. For any factor $R_A = 0.3838 < 0.493$ and $R_B = 0.3909 < 0.493$. Therefore, the equal variance assumption cannot be rejected on this basis.

Tables 4–6 present the results of the analysis for the variables X, Y, Z considering $\alpha=0.05$. For the three sources of variation (factor A, B and interaction AB) we provide sum squares (SS), degrees of freedom (df), mean square (MS), and F, the value for the statistical test. The computed tests for the three sources are expressed as

$$F(A)=\frac{\mathrm{MS}(A)}{\mathrm{MS}(AB)}; \quad F(B)=\frac{\mathrm{MS}(B)}{\mathrm{MS}(AB)}; \quad F(AB)=\frac{\mathrm{MS}(AB)}{\mathrm{MS}(\mathrm{Error})}. \tag{15}$$

For the level of significance mentioned it is required the next critical values: $F_{AB} \in [2.45-2.53]$, $F_A = 6.94$ and $F_B = 6.94$. By comparing the test statistics with the critical values of the F distributions, we find that: (1) variables X and Y are not influenced by the studied factors, (2) we can assume that there is not interaction and (3) variable Z is influenced by the studied factors, in this case the variance between treatments is significantly greater than the variance within treatment. Then, the null hypotheses of (12) and (3), $\sigma_{Ai}^2=0$ and $\sigma_{Bj}^2=0$, respectively, are rejected on this basis.

Table 4
Statistical analysis of variable X

Source of variation		SS	df	MS	F
A	Hypothesis	175.231	2	87.616	0.084
	Error	4159.578	4	1039.895	
B	Hypothesis	819.899	2	409.949	0.394
	Error	4159.578	4	1039.895	
AB interaction	Hypothesis	4159.578	4	1039.895	0.450
	Error	208,119.616	90	2312.440	

Table 5
Statistical analysis of variable Y

Source of variation		SS	df	MS	F
A	Hypothesis	524.886	2	262.443	0.807
	Error	1301.307	4	325.327	
B	Hypothesis	243.306	2	121.653	0.374
	Error	1301.307	4	325.327	
AB interaction	Hypothesis	1301.307	4	325.327	0.663
	Error	46,229.843	90	513.665	

Table 6
Statistical analysis of variable Z

Source of variation		SS	df	MS	F
A	Hypothesis	625.430	2	312.715	14.626
	Error	85.525	4	21.381	
B	Hypothesis	1966.444	2	983.222	45.985
	Error	85.525	4	21.381	
AB interaction	Hypothesis	85.525	4	21.381	0.038
	Error	50,978.605	90	566.429	

In Figs. 1–3 the horizontal axis represents the different number of observations (factor A) and the vertical axis represents the different time interval (factor B).

Figs. 1 and 2 show that the three graphed lines are not far away, with significance, from a horizontal line. On the other hand, in Fig. 3 for the variable Z the lines are different, they are separated and are not horizontal with significance.

Rejecting the null hypothesis for the variable Z suggests that treatment means differ, but does not signify where such differences lie. Descriptive comparisons are helpful in delineating where

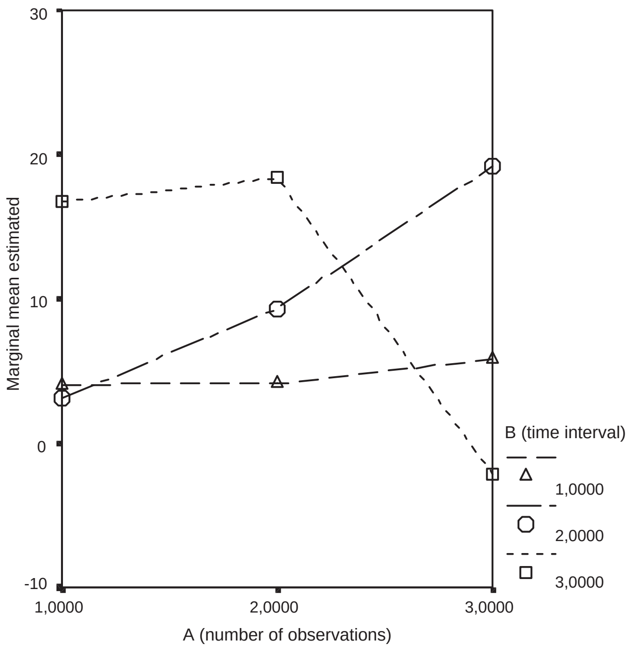

Fig. 1. Marginal mean estimated for X.

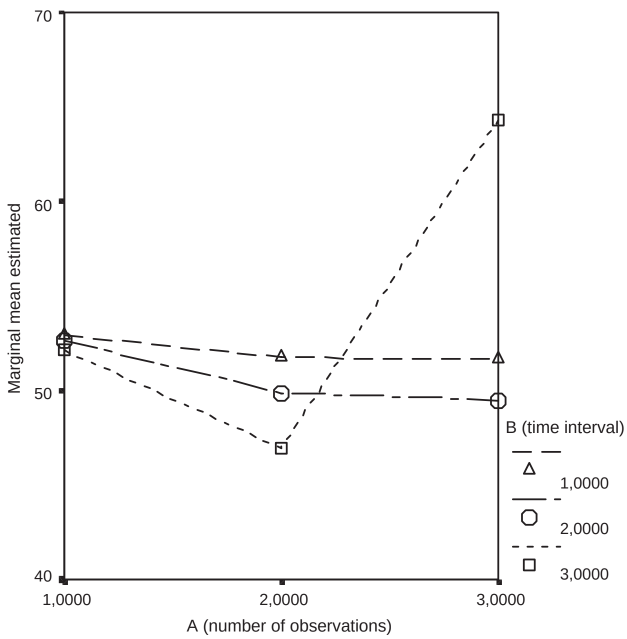

Fig. 2. Marginal mean estimated for variable Y.

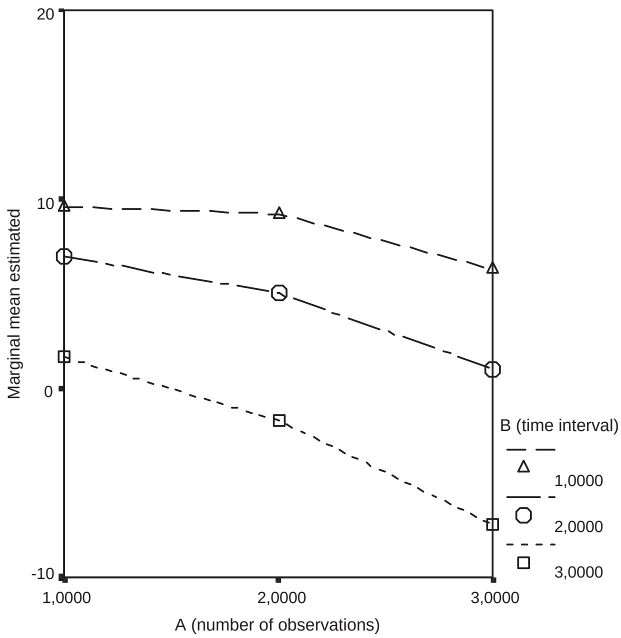

Fig. 3. Marginal mean estimated for variable Z.

Table 7
HSD test for multiple comparisons

Factor	$\mu_1 - \mu_2$	$\mu_1 - \mu_3$	$\mu_2 - \mu_3$
A (number of observations)	2.65	6.78	4.12
B (time interval)	4.04	10.80	6.76

differences lie, but are inconclusive in their results. Therefore, post hoc multiple comparisons are needed. The *Honestly Significant difference Tukey test* (HSD) enable us to find where the differences lie by pursing multiple tests. Then the null hypotheses for both factors are (1) $\mu_1 = \mu_2$, (2) $\mu_1 = \mu_3$ and (3) $\mu_2 = \mu_3$.

The HSD test reveals that for a level of significance of 0.05, HSD = 2.72. In Table 7 it is observed that are five values which are greater than HSD, therefore the null hypotheses are rejected. In completing all six tests using HSD we find that comparison for factor A, μ_1 vs. μ_2 is not significant. All other comparisons are significant.

5. Discussions and conclusions

In this paper, we provide a statistical explanation of the greater inaccuracy of the vertical coordinate relative to the two horizontal coordinates arising in the resolution of GPS equations. From the statistical study on this paper we can conclude that the X and Y solution are independent of the election of the time interval between observations and the number of observations factors. However,

Z unknown shows a significant effect in the manner that time interval and the number of observations are chosen. We next summarize that result:

(1) (Variable X), $\sigma^2_{Bj} = \sigma^2_{Ai} = 0$, hypothesis accepted,
(2) (Variable Y), $\sigma^2_{Bj} = \sigma^2_{Ai} = 0$, hypothesis accepted,
(3) (Variable Z), $\sigma^2_{Bj} = \sigma^2_{Ai} = 0$, hypothesis rejected,
(4) (Variable X, Y, Z), $\sigma^2_{ABij} = 0$, hypothesis accepted.

Although the three coordinates X, Y and Z represent the same kind of unknown (station position), we statistically found that parameter Z is not independent of the elected equation system.

This seems to be in concordance with numerical examples derived from the fundamental error equation in GPS error analysis and the perturbation analysis by the implicit function theorem as pointed out in Section 2. In this case, the greater error observed for the Z coordinate may indicate unstableness of Z caused by variation in both time and geometry, as our statistical study shows here. When we vary the number of equations (factor A) it is included new time unknowns and new geometry configurations. Moreover, when we vary the time interval between observations (factor B), it is changed the geometry configuration. It can be understood as perturbation on times in (7) and perturbation on geometry in (9).

Additionally, large correlations has been experimentally tested between vertical errors and timing errors in general, see [4,5]. This fact also may support our basis of the greater inaccuracy of the vertical coordinate relative to the horizontal coordinates when times is perturbed.

A further study on the covariances matrices from the least-squares method may provide qualitative information about the differences among the elected equation system.

Acknowledgements

The authors thank the anonymous referees for their helpful suggestions and comments.

References

[1] P. Heroux, J. Kouba, GPS precise Point Positioning with a difference, Geomatics'95, Ottawa, Canada, 1995.
[2] J. Neter, W. Wasserman, M.H. Kutner, Applied Linear Statistical Models, Richard D. Irwin Inc., Homewood, IL, 1985.
[3] G. Nord, D. Jabon, J. Nord, The global positioning system and the implicit function theorem, SIAM Rev. 40 (3) (1998) 692–696.
[4] M.C. Olynik, Temporal characteristics of GPS error source and their impact on relative positioning, M.Sc. Thesis, Department of Geomatics Engineering, Calgary, Alberta, July 2002.
[5] B.W. Parkinson, in: B.W. Parkinson, J.J. Spilker (Eds.), GPS Error analysis, in Global Positioning System: Theory and Applications, Vol. I, American Institute of Aeronautics and Astronautics Inc., New York, 1996.
[6] G. Strang, K. Borre, Linear Algebra, Geodesy, and GPS, Wellesley-Cambridge Press, Wellesley, 1997.